**House Price Prediction Using Boston Housing Dataset**

# Objective

This project involves predicting house prices using the Boston Housing Dataset. We implemented three regression models from scratch – Linear Regression, Random Forest, and XGBoost – to evaluate their performance and interpret their predictions.

# Methodology

1. **Data Preprocessing:** The dataset was normalized and cleaned. It was loaded manually due to the deprecation of `load\_boston` from Scikit-learn.
2. **Model Implementation:** Linear Regression, Random Forest, and XGBoost were implemented from scratch, avoiding high-level libraries like `sklearn.linear\_model`.
3. **Model Evaluation:** Models were compared using RMSE and R² metrics.
4. **Feature Importance:** Feature importance for Random Forest and XGBoost was visualized to understand the contribution of each feature.

# Problems Faced and Solutions

1. The original Boston dataset was removed from Scikit-learn due to ethical concerns. To overcome this, the dataset was sourced manually from Kaggle.
2. Feature importance was initially missing for the custom tree-based models. We resolved this by averaging feature importances from all individual decision trees.
3. File upload issues were encountered in Colab, and we used manual upload as a workaround.

# Findings

* Among the three models, XGBoost performed the best in terms of both RMSE and R², indicating higher prediction accuracy.
* Feature importance visualizations helped identify that features like 'LSTAT', 'RM', and 'PTRATIO' had significant influence on house prices.
* Random Forest also performed well, offering interpretable predictions, while Linear Regression had higher error due to its simplistic assumptions.

# Conclusion

This project highlighted how custom model implementations can help understand internal mechanics and gain interpretability. Despite challenges in dataset access and feature importance extraction, we were able to successfully build, evaluate, and interpret models for house price prediction.